PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-041078

(43)Date of publication of application : 08.02.2002

(51)Int.Cl. G10L 15/06
G10L 15/10
G10L 15/00
G10L 15/28

(21)Application number : 2000-220576

(22)Date of filing : 21.07.2000

(71)Applicant : SHARP CORP

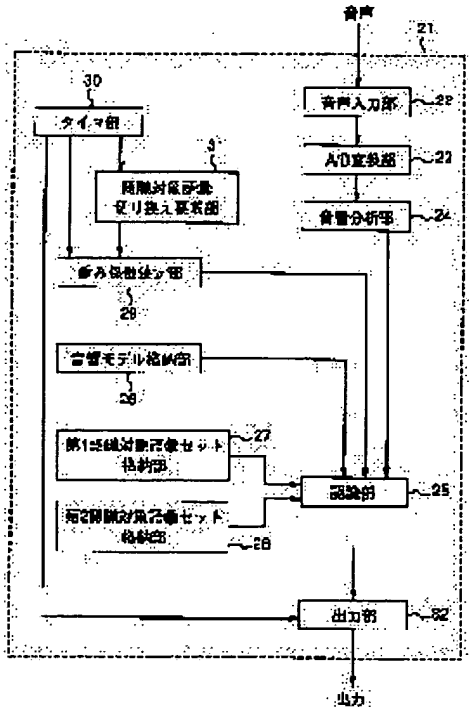
(72)Inventor : HONDA KAZUMASA
TSURUTA AKIRA
KANZA HIROYUKI

(54) VOICE RECOGNITION EQUIPMENT, VOICE RECOGNITION METHOD AND PROGRAM RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain high recognition accuracy even when the recognition object vocabulary is automatically changed.

SOLUTION: A recognition means 25 calculates the likelihood P of the vocabulary which constitutes the recognition object vocabulary sets A, B stored in the first and second recognition object vocabulary storing parts 27, 28 using the acoustics model in an acoustics model storing part 26. The change of the recognition object vocabulary sets accompanied with the display contents in an output part 32 is performed at the time by changing the values of weights w1, w2 which multiply to P between '1' and the appointed value 'a' near 0 (zero) proportioned the passing time from the requested change time t0. As the result, even when the speaker misses the utterance chance and the recognition object vocabulary is changed automatically, high recognition result can be obtained if the speaker utters the recognition object vocabulary before changing because the calculations of the likelihood w×P are also performed.



LEGAL STATUS

[Date of request for examination] 26.07.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]	3563018
[Date of registration]	11.06.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

CLAIMS

[Claim(s)]
[Claim 1] The recognition section which recognizes the inputted voice, and the output section which outputs information including the recognition result of this recognition section, The recognition lexical storing section in which the vocabulary for recognition used at the time of the above-mentioned recognition was stored, In the voice recognition unit which has the timer section and the lexical switch demand section for recognition which requires a switch of the above-mentioned vocabulary for recognition based on the time-of-day signal from this timer section the above-mentioned output section The switch output of two or more contents of an output is carried out. The above-mentioned vocabulary for recognition It is classified into two or more lexical sets for recognition which become by the set of the word for recognition corresponding to the content of an output of the above-mentioned output section. A switch of the above-mentioned vocabulary for recognition is performed in the unit of the above-mentioned lexical set for recognition. It is the voice recognition unit which is equipped with the weight decision section which determines the weight for each above-mentioned lexical set for recognition based on the time-of-day signal from the above-mentioned timer section, and is characterized by the above-mentioned recognition section recognizing input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].
[Claim 2] It is the voice recognition unit carried out [that it raises the weight for the lexical set for recognition after a switch in a voice recognition unit according to claim 1 while the above-mentioned weight decision section reduces the weight for the lexical set for recognition before a switch according to the elapsed time to weight decision, after a switch of the vocabulary for recognition is required by the above-mentioned lexical switch demand section for recognition, and] as the description.
[Claim 3] It is the voice recognition unit characterized by computing the likelihood of each word from which the above-mentioned recognition section constitutes all the above-mentioned lexical sets for recognition in claim 1 or a voice recognition unit according to claim 2, applying the weight for the lexical set for recognition with which each word belongs to the value of the likelihood of each word, and making a word with the highest value into a recognition result.
[Claim 4] In a voice recognition unit according to claim 2 the above-mentioned output section The value of the weight for the lexical set for recognition corresponding to the content of an output currently outputted when the lexical switch demand for recognition from the above-mentioned lexical switch demand section for recognition is made, Next, the voice recognition unit which will be characterized by switching the above-mentioned content of an output if a difference with the value of the weight for the lexical set for recognition corresponding to the content of an output which should be outputted becomes under a predetermined value.
[Claim 5] In the speech recognition approach which faces recognizing the inputted voice using the vocabulary for recognition, and outputting a recognition result, and switches the above-mentioned vocabulary for recognition automatically based on the time-of-day signal from the timer section Two or more contents of an output in the unit of two or more lexical sets for recognition which carry out a switch output and become the output section by the set of the word for recognition corresponding to each above-mentioned content of an output The speech recognition approach characterized by switching the above-mentioned vocabulary for recognition, determining the weight for each above-mentioned lexical set for recognition based on the time-of-day signal from the above-mentioned timer section, and recognizing the above-mentioned input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].
[Claim 6] The program documentation medium which is characterized by recording the speech recognition processing program as which a computer is operated as the recognition section in claim 1, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section and in which computer read-out is possible.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the program documentation medium which recorded the speech recognition processing program on the voice recognition unit and the speech recognition approach of being carried in a computer or a Personal Digital Assistant and recognizing the voice by utterance of human being, and the list.
[0002]
[Description of the Prior Art] In a voice recognition unit, in order to raise recognition precision, there is the recognition approach of switching the vocabulary for recognition if needed. As an application of the voice recognition unit using such a recognition approach, it is possible to perform the actuation guide of the device by the menu display using a display using speech recognition in the device which has displays, such as a personal computer and a Japanese word processor.
[0003] According to the above actuation guides, actuation can be studied, checking the display of the effectiveness by operating instructions or actuation in the pictures. And when there is little amount of information from the above-mentioned display -- the screen of the above-mentioned display is narrow -- the display of the actuation guide about two or more device actuation may be automatically switched with the passage of time. If voice is used for such an actuation guide, for a user, it is intelligible, and the number of manual operation buttons can be reduced and actuation can be simplified. In that case, if the vocabulary for recognition is switched with a switch of a display of the actuation guide about two or more device actuation, since the vocabulary for recognition can be lessened, a high recognition precision can be acquired.
[0004] In application of the recognition approach which switches such a vocabulary for recognition, two or more sets of the vocabulary for recognition which has relation in each menu which indicates by switch are memorized only several menu minutes. And by switching the vocabulary for recognition synchronizing with a switch of the menu display by actuation of a user, the passage of time, etc., in each menu, recognition processing can be performed for a necessary minimum vocabulary, and recognition precision can be raised. In that case, in case a menu display is automatically switched with the passage of time, a device will also switch the vocabulary for recognition automatically.
[0005] Hereafter, the voice recognition unit which can switch the above-mentioned vocabulary for recognition is explained. Drawing 4 is the block diagram showing an example of the voice recognition unit which can switch the above-mentioned vocabulary for recognition. Here, voice recognition unit 1 self shall perform automatically a switch of the content of a display according [this voice recognition unit 1] to a switch of the vocabulary for recognition, and the output section 13 for every predetermined time. A voice recognition unit 1 consists of the A/D (analog/digital) converter 2, the sonagraphy section 3, the recognition section 4, the sound model storing section 5, lexical storing [for recognition] / judgment section 6, the lexical identifier storage section 7 for the present recognition, the timer section 8, the lexical switch demand section 9 for recognition, the lexical switch demand time-of-day storage section 10 for recognition, a voice detecting element 11, the voice time-of-day storage section 12, and the output section 13.
[0006] The voice inputted into the above-mentioned voice recognition unit 1 is sent out and digitized by the A/D-conversion section 2 by the speaker. And this digitized voice wave is the sonagraphy section 3, and it is analyzed by the technique of the short-time analysis of a spectrum which shifts the above-mentioned time window every 8msec - 16msec while it hangs a short-time time window comparatively for every section of 20msec - 40msec. The voice wave started by the above-mentioned time window is changed into the time series of the feature vector of the unit called the frame which has the time amount length at the time of logging. Here, the above-mentioned feature vector is what extracted ****** of the voice spectrum in the time of day, it is usually 10-dimensional one - 100 dimensions, and the LPC (linear predictive coding) mel cepstrum multiplier etc. is used widely. In this way, the changed feature vector is outputted also to the voice detecting element 11 which detects initiation of voice input while it is sent out to the recognition section 4. If it does so, the voice time-of-day storage section 12 will detect and memorize the start time of voice input based on the voice input start signal from the voice detecting element 11, and the time-of-day signal from the timer section 8.
[0007] HMM (hidden Markov model) prepared for every recognition unit is prepared for the above-mentioned sound model storing section 5. Here, as the above-mentioned recognition unit, the phoneme and the word are used widely. Moreover, in HMM, it is the nondeterminism probability finite automaton which has two or more conditions, and is the source model of a statistical signal which expresses the source of an unsteady signal with connection of the source of a normal signal. In addition, parameters, such as a output probability and transition probability, are beforehand learned by the algorithm which gives corresponding study voice and is called a BAUMU-Welch algorithm. Hereafter, HMM whose recognition unit is a phoneme shall be memorized by the sound model storing section 5.
[0008] Actuation of a switch of the above-mentioned vocabulary for recognition applies the approach currently indicated by JP,6-337695,A. As the above-mentioned vocabulary for recognition, there shall be a lexical set A for recognition and a lexical set B for recognition, and the identifier of the lexical set A for recognition shall be memorized in this time by the lexical identifier storage section 7 for recognition. Moreover, the output section 13 shall show the content of a display corresponding to the lexical set A for recognition.
[0009] In this condition, progress of predetermined time makes advice from the timer section 8 to the lexical switch demand

section 9 for recognition, and the output section 13. If it does so, the output section 13 will change the content of a display into the content of a display corresponding to the lexical set B for recognition. Moreover, a switch is required from the lexical switch demand section 9 for recognition, and the demand time of day is memorized by the lexical switch demand time-of-day storage section 10 for recognition. And the voice input start time Ts memorized by lexical storing [for recognition] / judgment section 6 at the demand time of day Tc and the voice time-of-day storage section 12 which are memorized by the lexical switch demand time-of-day storage section 10 for recognition is compared. When the voice input start time Ts is the back [time of day / Tc / demand], since utterance was performed after the switch of the vocabulary for recognition was required, it is judged with the suitable lexical set for recognition being the lexical set B for recognition. It is judged with it being the lexical set A for recognition except it. And the content of storage of the lexical identifier storage section 7 for the present recognition is updated by the identifier of the corresponding lexical set for recognition.

[0010] After the judgment of the suitable lexical set for recognition is completed, in this way, the recognition section 4 The phoneme train of each word which constitutes which lexical set for recognition outputted from lexical storing [for recognition] / judgment section 6 corresponding to the feature vector obtained in the sonagraphy section 3, and the identifier memorized by the lexical identifier storage section 7 for the present recognition, Speech recognition is performed as follows using HMM stored in the sound model storing section 5.

[0011] That is, HMM of each word contained in the above-mentioned vocabulary for recognition is calculated first. HMM of each phoneme memorized by the sound model storing section 5 is made to specifically correspond to the phoneme train of each word which constitutes the lexical set for recognition, and it joins together.

[0012] Next, an occurrence probability is searched for about HMM of each word using the feature vector from the sonagraphy section 3. In the speech recognition by HMM, voice is expressed as time series of the symbol outputted from HMM between the state transitions from an initial state to a final state. Then, the probability for utterance to be generated from the model M (HMM of a word) can be searched for by setting the probability of an initial state to any value, and imposing a output probability and transition probability for every state transition one by one. On the contrary, when utterance is observed and it assumes that it generated from the model M with the utterance, the probability of generating from the model M can be calculated.

[0013] Hereafter, the recognition algorithm in the above-mentioned recognition section 4 is explained to a detail. The recognition section 4 considers the time series of the feature vector obtained by the sonagraphy section 3 as an input, searches for the occurrence probability about HMM of all the words contained in the vocabulary for recognition from lexical storing [for recognition] / judgment section 6, and makes a recognition result the word of HMM which presents the highest occurrence probability. Namely, the sequence of the input expressed by the time series of a feature vector is set to X=xvec1, xvec2 and xvec3, --, xvect, --, xvecI by making t (= 1, 2, --, I) into a frame number. In addition, "xveci" is the vector of many dimensions. Hereafter, "xvec" is written. [Vector x] Furthermore, the set of the initial state of Model M is set to S, and the set of a final state is set to F. Moreover, the sere of the j-th condition is expressed as Q=qOj, q1j, q2j, --, qtj, --, qIj by making "i, j" into a state number. In a top type, "qtj" expresses the condition of having changed by the input symbol xvect of the t-th frame. Here, it is qOj**S and qIj**F. Furthermore, when the initial probability of an initial state is expressed with pi i:sigma qi**Spii =1, transition probability from Condition qi to Condition qj is set to aij and the output probability by which xveci is then outputted is set to bij (xveci), occurrence probability (likelihood) P(X|M) of an input sequence is,

$$P(X \mid M) = \sum_{allQj} \pi_0^j \prod_{i=1}^{I} a_{t-1,t}^j \cdot b_{t-1,t}^j (x_{vec} i)$$

It is come out and expressed. It calculates by HMM-attaching, and outputs and displays on the output section 13 by making into a recognition result the word corresponding to HMM corresponding to all the words contained in the vocabulary for recognition in the operation of this occurrence probability (likelihood) P(X|M) which presents the highest occurrence probability (likelihood) P.

[0014]

[Problem(s) to be Solved by the Invention] However, there are the following problems in the voice recognition unit which applied the lexical switch actuation for recognition indicated by above-mentioned conventional JP,6-337695,A. That is, as mentioned above, when the voice input start time Ts is the back [time of day / Tc / for recognition / lexical switch demand] in the lexical switch actuation for recognition indicated by JP,6-337695,A, he is trying to switch the set of the vocabulary for recognition. When the lexical switch demand for recognition is made by actuation of a speaker, since utterance is performed after the switch demand of the vocabulary for recognition is surely made, this approach is effective.

[0015] However, like the voice recognition unit shown in drawing 4 , when it is the voice recognition unit which the vocabulary for recognition turns off and is automatically replaced with the passage of time, a switch of the vocabulary for recognition is performed regardless of a speaker's consciousness at all. Therefore, a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and when a switch of the vocabulary for recognition is performed automatically, the need of returning to the established state of the vocabulary for recognition before the switch which the speaker wanted to utter by a certain approach arises. And there is a problem that a speaker will be kept waiting until it makes a speaker pay a certain actuation or the vocabulary for recognition before a switch is set up automatically in that case.

[0016] Then, the object of this invention is to offer the program documentation medium which recorded the speech recognition processing program on the voice recognition unit which is easy to use and the speech recognition approach that a high recognition precision is acquired, and the list, even when switching the vocabulary for recognition automatically.

[0017]

[Means for Solving the Problem] The recognition section which recognizes the voice into which the 1st invention was inputted in order to attain the above-mentioned object, The output section which outputs information including the recognition result of this recognition section, and the recognition lexical storing section in which the vocabulary for recognition used at the time of the above-mentioned recognition was stored, In the voice recognition unit which has the timer section and the lexical switch demand section for recognition which requires a switch of the above-mentioned vocabulary for recognition based on the time-of-day signal from this timer section the above-mentioned output section The switch output of two or more contents

of an output is carried out. The above-mentioned vocabulary for recognition It is classified into two or more lexical sets for recognition which become by the set of the word for recognition corresponding to the content of an output of the above-mentioned output section, and a switch of the above-mentioned vocabulary for recognition is performed in the unit of the above-mentioned lexical set for recognition. Based on the time-of-day signal from the above-mentioned timer section, it has the weight decision section which determines the weight for each above-mentioned lexical set for recognition, and the above-mentioned recognition section is characterized by recognizing input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].

[0018] According to the above-mentioned configuration, input voice is recognized by the recognition section using the weight for each lexical set for recognition determined by the weight decision section based on the time-of-day signal from all the lexical sets for recognition, and the timer section. If a switch of the vocabulary for recognition is required by the lexical switch demand section for recognition based on the time-of-day signal from the above-mentioned timer section in that case, the lexical set for recognition used now will be switched to the lexical set for recognition according to a switch of the content of an output of the output section. Therefore, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised.

[0019] Furthermore, since recognition is performed also using the word of the lexical set for recognition before a switch even if it utters with the vocabulary for recognition before a switch to not knowing, a high recognition precision is acquired [that the above-mentioned lexical set for recognition was switched for the speaker, and] also about the word of the lexical set for recognition before the above-mentioned switch.

[0020] Moreover, after the above-mentioned weight decision section is required of a switch of the vocabulary for recognition by the above-mentioned lexical switch demand section for recognition, while the voice recognition unit [above / 1st] of invention reduces the weight for the lexical set for recognition before a switch according to the elapsed time to weight decision, it is [voice recognition unit] desirable in accomplishing so that the weight for the lexical set for recognition after a switch may be raised.

[0021] While according to the above-mentioned configuration the elapsed time after a switch of the vocabulary for recognition is required by the above-mentioned lexical switch demand section for recognition takes for becoming long and the recognition precision of the vocabulary for recognition before a switch becomes low, the recognition precision of the vocabulary for recognition after a switch becomes high. In this way, a switch of the above-mentioned vocabulary for recognition used for recognition is performed gradually.

[0022] Moreover, as for the voice recognition unit of invention of the above 1st, it is desirable to accomplish so that the likelihood of each word which constitutes all the above-mentioned lexical sets for recognition for the above-mentioned recognition section may be computed, the weight for the lexical set for recognition with which each word belongs to the value of the likelihood of each word may be applied and a word with the highest value may be made into a recognition result.

[0023] Obtaining raising the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section and a recognition precision high even when a speaker utters at the vocabulary for recognition before switching is easily attained by setting up the weight for the lexical set for recognition used for recognition, and the weight for the lexical set for recognition which is not used for recognition the optimal according to the above-mentioned configuration.

[0024] Moreover, when the difference of the value of the weight for the lexical set for recognition corresponding to the content of an output which is outputting the above-mentioned output section when the lexical switch demand for recognition from the above-mentioned lexical switch demand section for recognition is made, and the value of the weight for the lexical set for recognition corresponding to the content of an output which should output to a degree becomes under in a predetermined value, it is [voice recognition unit / of invention of the above 1st] desirable in accomplishing so that the above-mentioned content of an output may switch.

[0025] According to the above-mentioned configuration, in response to the above-mentioned lexical set for recognition being switched, it is switched to the content of an output to which the content of an output of the above-mentioned output section is equivalent.

[0026] Moreover, face the speech recognition approach of the 2nd invention recognizing the inputted voice using the vocabulary for recognition, and outputting a recognition result, and it is set to the speech recognition approach which switches the above-mentioned vocabulary for recognition automatically based on the time-of-day signal from the timer section. Carry out the switch output of two or more contents of an output at the output section, and the above-mentioned vocabulary for recognition is switched in the unit of two or more lexical sets for recognition which become by the set of the word for recognition corresponding to each above-mentioned content of an output. Based on the time-of-day signal from the above-mentioned timer section, the weight for each above-mentioned lexical set for recognition is determined, and it is characterized by recognizing the above-mentioned input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].

[0027] According to the above-mentioned configuration, input voice is recognized using the weight for each lexical set for recognition determined based on the time-of-day signal from all the lexical sets for recognition, and the timer section. If a switch of the vocabulary for recognition is required based on the time-of-day signal from the above-mentioned timer section in that case, the lexical set for recognition used now will be switched to the lexical set for recognition according to a switch of the content of an output of the output section. Therefore, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised.

[0028] Furthermore, since recognition is performed also using the word of the lexical set for recognition before a switch even if it utters with the vocabulary for recognition before a switch to not knowing, a high recognition precision is acquired [that the above-mentioned lexical set for recognition was switched for the speaker, and] also about the word of the lexical set for recognition before the above-mentioned switch.

[0029] Moreover, the program documentation medium of the 3rd invention is characterized by recording the speech

recognition processing program as which a computer is operated as the recognition section in claim 1, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section.

[0030] According to the above-mentioned configuration, like the case of claim 1, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised. Furthermore, even if a speaker utters that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before switching to not knowing, a high recognition precision is acquired.

[0031]

[Embodiment of the Invention] Hereafter, the gestalt of implementation of a graphic display explains this invention to a detail. Drawing 1 is a block diagram in the voice recognition unit of the gestalt of this operation. This voice recognition unit 21 consists of the voice input section 22, the A/D-conversion section 23, the sonagraphy section 24, the recognition section 25, the sound model storing section 26, the lexical set storing section 27 for the 1st recognition, the lexical set storing section 28 for the 2nd recognition, the weighting-factor decision section 29, the timer section 30, the lexical switch demand section 31 for recognition, and the output section 32.

[0032] The above-mentioned voice input section 22 is equipped with the audio input unit containing a microphone, changes the inputted voice into an electrical signal (sound signal), and outputs it to the A/D-conversion section 23. The A/D-conversion section 23 changes the sound signal which is an inputted analog signal into a digital signal, and outputs the digitized sound signal to the sonagraphy section 24. In addition, the sound signal by which digitization was carried out [above-mentioned] is expressed with the time series of amplitude value.

[0033] The above-mentioned sonagraphy section 24 extracts a feature vector from the digital sound signal from the A/D-conversion section 23 for every frame, and outputs it to the recognition section 25. Here, the above-mentioned feature vector continues and arranges the 34-dimensional vector xvec which consists of power of the power of the sound signal in each frame, primary - the 16th LPC cepstrum multiplier, and a before frame, and a total of 34 elements of the LPC cepstrum multiplier (primary - 16th order) of a before frame on all frames (t= 1, 2, --, I).

[0034] The above-mentioned recognition section 25 calculates the occurrence probability (likelihood) P of each word which constitutes the lexical set B for recognition stored in the lexical set A for recognition stored in the lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition using a sound model using the technique explained by the Prior art using the feature vector extracted in the sonagraphy section 24. Furthermore, weight w determined in the weighting-factor decision section 29 is hung on the likelihood P of each word, and the word corresponding to HMM which presents highest likelihood w-P is outputted to the output section 32.

[0035] The sound model used in case the above-mentioned sound model storing section 26 performs speech recognition in the recognition section 25 is stored. The above-mentioned sound model makes a phoneme a unit, and HMM learned by the algorithm beforehand called a BAUMU-Welch algorithm using the study voice of an unspecified speaker (initial study) is used. In addition, Above HMM is memorized in the array for several condition minutes which uses the transition probability and output probability distribution in each condition as an element. Moreover, the above-mentioned transition probability is memorized in the array for several transition minutes by using transition probability to each condition as an element. Moreover, the above-mentioned output probability is expressed with the contaminated normal distribution of the many dimensions which carried out weighting addition of two or more normal distribution, and is memorized in the array for the number of dimension which uses the weight, the mixed mean vector, and mixed distributed vector in each normal distribution as an element. Here, it is expressed with the array of the element of "34" as the number of elements of the feature vector extracted from a sound signal for every frame in the sonagraphy section 24 with same above-mentioned mean vector and distributed vector.

[0036] The above-mentioned timer section 30 outputs the time-of-day signal showing time of day to the lexical switch demand section 31 for recognition, the weighting-factor decision section 29, and the output section 32, and notifies time of day. If it does so, the lexical switch demand section 31 for recognition will judge whether a switch of the vocabulary for recognition is required based on the time of day by which advice was given [above-mentioned]. And in requiring, it requires a switch of the vocabulary for recognition from the weighting-factor decision section 29.

[0037] The inside of the lexical set B for recognition stored in the lexical set A for recognition with which the above-mentioned weighting-factor decision section 29 is stored in the lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition, The weight w2 applied to the word which constitutes the lexical set for recognition corresponding to the content of a display in which it is indicated by current by the output section 32, and the weight w1 applied to the word which constitutes the lexical set for recognition corresponding to the content of a display which is not shown by the output section 32 are determined. Such weight w1 and w2 is determined whenever predetermined time deltaT passes using the weight function W1 (t) memorized and W2 (t) on the basis of the time of day tO from the timer section 30 when a switch is required from the lexical switch demand section 31 for recognition. And the sequential output of the value of both the determined weight w1 and w2 is carried out at the recognition section 25.

[0038] The word which constitutes each lexical set A and B for recognition is memorized by the above-mentioned lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition in the array for several alphabetic character minutes which uses the character string of the notation of each word, and a phoneme train as an element.

[0039] The above-mentioned output section 32 is equipped with an image display device including a display, and stores the content of the 1st display corresponding to the lexical set A for recognition, and the content of the 2nd display corresponding to the lexical set B for recognition. And it judges whether based on the time of day notified from the timer section 30, the content of a display which is indicating by current among the contents of the 1st and 2nd display is changed, and when changing, the content of a display of a screen is switched. Furthermore, the recognition result from the recognition section 25 is displayed on a screen.

[0040] Drawing 2 shows time amount change with the weight function W2 for the lexical set for recognition corresponding to the content of a display which the above-mentioned output section 32 has chosen now (t), and the weight function W1 for the

lexical set for recognition corresponding to the content of a non-choosing display (t). It is begun from about zero predetermined value "a" smaller than 1 to carry out the monotonous increment of the value of a weight function W1 (t) at the time of day t0 when the switch demand of the vocabulary for recognition was outputted, and becomes a value "1" after time of day t2. On the other hand, it is begun at time of day t0 to carry out monotonous reduction of the value of a weight function W2 (t) with the value of a weight function W1 (t) at reverse from a value "1", and becomes a predetermined value "a" after time of day t2. In that case, the difference of weight w1 and weight w2 serves as a threshold h at time of day t1. And the output section 32 will switch the content of a display currently displayed on the screen, if the value of this difference turns into under the threshold h (i.e., if time amount T (> (t1-t0)) passes since the time of day t0 when the switch of the vocabulary for recognition was demanded).

[0041] That is, when judging that the above-mentioned output section 32 changes the content of a display based on the time of day notified from the timer section 30, it is set up so that only the above-mentioned time amount T may be late for the event of the lexical switch demand section 31 for recognition judging that the above-mentioned switch is required based on the time of day notified from the timer section 30.

[0042] Thus, in the gestalt of this operation, although the content of a display of a screen is automatically switched by the output section 32, even if it is after cutting and replacing, even if it is before the content of a display cuts and replaces, the recognition section 25 calculates likelihood P for the vocabulary of the lexical set A for recognition, and both the lexical set of the lexical set B for recognition. And if it is before the content switch of a display, it is 1>w> (1+a+h)/2, and if it is after a switch, weight w which is 1>w> (1+a-h)/2 will be hung on the likelihood P of the word which constitutes the lexical set for recognition corresponding to the content of a display chosen by the current output section 32. On the other hand, if it is before the content switch of a display (1+a-h), it is /2>w>a, and if it is after a switch (1+a+h), weight w which is /2>w>a will be hung on the likelihood P of the word which constitutes the lexical set for recognition corresponding to the content of a display by the side of un-choosing. In this way, he calculates final likelihood w-P and is trying to determine a recognition result.

[0043] If it puts in another way, a switch of the vocabulary for recognition in the conventional voice recognition unit shown in drawing 4 It sets in the gestalt of this operation to carrying out by switching the vocabulary for recognition itself used for the operation of likelihood P. The two-set vocabulary for recognition used for the operation of likelihood P is performed by changing gradually the value of weight w hung on likelihood P, without switching between "1" and about zero predetermined value "a."

[0044] Therefore, in the gestalt of this operation, even if a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and count of likelihood w-P about the word of the vocabulary for recognition before a switch will also be performed also after a switch of the vocabulary for recognition is performed automatically, and it utters with the vocabulary for recognition before a speaker switching, it becomes possible to recognize correctly. Moreover, the function which raises the recognition precision of the vocabulary corresponding to the content of a display of the output section 32 is not spoiled like the case where the vocabulary for recognition itself is switched like the voice recognition unit shown in drawing 4 in that case.

[0045] Drawing 3 is the flow chart of the weight decision processing actuation performed by the above-mentioned weighting-factor decision section 29. Hereafter, actuation of weight decision is explained according to drawing 3 . Here, the output section 32 sets to W2 (t) the weight function for the lexical set for recognition corresponding to the content of a display which is making current selection, and sets the weight function for the lexical set for recognition corresponding to the content of a non-choosing display to W1 (t). If a switch is required from the lexical switch demand section 31 for recognition, weight decision processing actuation will start.

[0046] At step S1, the switch demand time of day t0 of the vocabulary for recognition is acquired based on the time-of-day signal from the above-mentioned timer section 30. At step S2, the count j of calculation of the weight value w is initialized by "0." At step S3, the increment of the count j of calculation is carried out. By step S4, since the switch demand time of day t0 is acquired, or after computing the weight value w last time, it is distinguished whether predetermined time deltaT passed. Consequently, if it has passed, it will progress to step S5. At step S5, it is distinguished whether current time of day (t0+j-deltaT) is over time of day t2. Consequently, if it has exceeded, while ending weight decision processing actuation, if it has not exceeded, it progresses to step S6.

[0047] At step S6, the function number i of the above-mentioned weight function Wi (t) is initialized by "1." At step S7, the value wi of weight is computed by "j-deltaT" being substituted for elapsed time t from the switch demand time of day t0 in a weight function W1 (t). The increment of the function number i is carried out at step S8. By step S9, it is distinguished whether the value of the function number i is larger than "2." Consequently, if larger than "2" while with "2" carrying out a return to step S7 and shifting to calculation of the weight value w2, it will be judged that the weight in the present time of day corresponding to all the lexical sets A and B for recognition was computed, and it will progress to step S10. [below] At step S10, the array of the weight values w1 and w2 in the present time of day by which calculation was carried out [above-mentioned] is outputted to the recognition section 25. A return is carried out to step S3 such the back, and it shifts to calculation of the weight values w1 and w2 in the next time of day.

[0048] Henceforth, if the above-mentioned step S3 - step S10 were repeated, current time of day is over time of day t2 in step S5 and it will be distinguished, weight decision processing actuation will be ended. After that, "1" is outputted to every predetermined time deltaT as a weight value w2 for the lexical set for recognition corresponding to the content of a display, and a predetermined value "a" is outputted to every predetermined time deltaT as a weight value w1 for the lexical set for recognition corresponding to the content of a non-choosing display. And if a switch demand is outputted from the lexical switch demand section 31 for recognition next, the above-mentioned weight decision processing actuation will start.

[0049] As mentioned above, the recognition section 25 in the gestalt of this operation computes the likelihood P of the word which constitutes the lexical set A for recognition stored in the lexical set storing section 27 for the 1st recognition, and the lexical set B for recognition stored in the lexical set storing section 28 for the 2nd recognition using the sound model stored in the sound model storing section 26. A switch of the lexical set for recognition accompanying a switch of the content of a display of the output section 32 in that case is performed by switching the value of the weight w2 and w1 applied to the

likelihood P of the word which constitutes selection and the lexical set for non-choosing recognition to "1" and about zero predetermined value "a" rather than switching the lexical set for recognition itself. And in that case, in proportion to the elapsed time "j-deltaT" from the time of day t0 when the switch demand was made from the lexical switch demand section 31 for recognition, he value "a" Passes from a value "1" gradually rather than switching the value of weight w2 and w1 gradually, or is trying to switch to a value "1" from a value "a."

[0050] Therefore, since according to the gestalt of this operation count of likelihood w-P about the word of the lexical set for recognition before a switch is also performed even if the vocabulary for recognition will be switched automatically [a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and], even if it utters with the vocabulary for recognition before a speaker switching, it can recognize correctly. Moreover, the function which raises the recognition precision of the vocabulary for recognition corresponding to the content of a display of the output section 32 is not spoiled like the case where the vocabulary for recognition itself is switched in that case like the voice recognition unit shown in drawing 4 .

[0051] In addition, he is trying to switch linearly the weight function W2 for the lexical set for selection recognition (t), and the weight function W1 for the lexical set for recognition corresponding to the content of a non-choosing display (t) to a value "a", and "1" from a value "1", and "a" in the gestalt of the above-mentioned implementation in proportion to the elapsed time "j-deltaT" from the switch demand time of day t0 by the lexical switch demand section 31 for recognition. However, in this invention, a function W1 (t) and the configuration of W2 (t) are not limited to a straight line. While making it a curve, lowering the value of a function W1 (t) while raising the value of the function W2 by the switch time of day t1 of the content of a display (t), and lowering the value of the function W2 after the switch time of day t1 of the content of a display (t), the value of a function W1 (t) may be raised.

[0052] Moreover, whenever predetermined-time deltaT passes the above-mentioned weighting-factor decision section 29 on the basis of the switch demand time of day tO from the lexical switch demand section 31 for recognition, it constitutes so that the weight values w1 and w2 may be determined and it may output to the recognition section 25, and in the gestalt of the above-mentioned implementation, the recognition section 25 constitutes so that the weight values w1 and w2 inputted may be used if needed and recognition processing may be carried out. However, this invention is not limited to this, in case it recognizes the recognition section 25, it is constituted so that a weight decision demand may be advanced to the weighting-factor decision section 29, and if a weight decision demand is received, the weighting-factor decision section 29 will not interfere, even if it constitutes so that the elapsed time from the switch demand time of day tO by the lexical switch demand section 31 for recognition may be substituted and computed to a weight function Wi (t).

[0053] By the way, the function as the above-mentioned recognition section in a gestalt, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section of each above-mentioned implementation is realized by the speech recognition processing program recorded on the program documentation medium. The above-mentioned program documentation media in the gestalt of the above-mentioned implementation are program media which become by ROM (read only memory). Or you may be the program media by which reading appearance is equipped with and carried out to external auxiliary storage. In addition, the program read-out means which reads a speech recognition processing program from the above-mentioned program media in the case of which may have the configuration which carries out direct access to the above-mentioned program media and which is read to them, and may download it to the program storage area (not shown) prepared in RAM (random access memory), and you may have the configuration accessed and read to the above-mentioned program storage area. In addition, the download program for downloading from the above-mentioned program media to the above-mentioned program storage area of RAM shall be beforehand stored in the main frame.

[0054] With the above-mentioned program media, it is constituted disengageable a body side here. Magnetic disks, such as tape systems, such as a magnetic tape and a cassette tape, a floppy (trademark) disk, and a hard disk, CD(compact disk)-ROM and MO (optical MAG) disk, MD (mini disc), The disk system of optical disks, such as DVD (digital video disc), It is the medium including semiconductor memory systems, such as card systems, such as IC (integrated circuit) card and an optical card, a mask ROM, EPROM (ultraviolet-rays elimination mold ROM) and EEPROM (electric elimination mold ROM), and a flash ROM, which **** a program fixed.

[0055] Moreover, if it has the configuration which the voice recognition unit in the gestalt of each above-mentioned implementation is equipped with a modem, and contains the Internet and in which a communication network and connection are possible, even if the above-mentioned program media are media which **** a program fluidly by download from a communication network etc., they will not interfere. In addition, the download program for downloading from the above-mentioned communication network which can be set in that case shall be beforehand stored in the main frame. Or it shall be installed from another record medium.

[0056] In addition, it is not limited only to a program and what is recorded on the above-mentioned record medium can also record data.

[0057]

[Effect of the Invention] So that clearly as mentioned above, the voice recognition unit of the 1st invention Two or more lexical sets for recognition corresponding to the content of an output of the output section are stored in the recognition lexical storing section. By the weight decision section Since the weight for each above-mentioned lexical set for recognition is determined based on the time-of-day signal from the timer section and input voice is recognized by the recognition section using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned] It is based on the switch demand of the vocabulary for recognition by the lexical switch demand section for recognition. If the value of the weight for the lexical set for recognition before a switch is lowered in case it is switched to the lexical set for recognition according to a switch of the content of an output of the above-mentioned output section, the recognition precision of the vocabulary for recognition after a switch can be raised.

[0058] Furthermore, since a speaker recognizes it to it also using the word of the lexical set for recognition before a switch even if he utters to not knowing that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before a switch, a high recognition precision can be acquired also about the word of the lexical set for recognition

before the above-mentioned switch.

[0059] That is, according to this invention, even when switching the vocabulary for recognition automatically, a high recognition precision can be acquired. Furthermore, the voice recognition unit which a speaker is not made to pay a certain actuation and latency time in that case, and is easy to use is realizable.

[0060] Moreover, the voice recognition unit [above / 1st] of invention can switch gradually the above-mentioned vocabulary for recognition used for recognition, if it accomplishes so that the weight for the lexical set for recognition after a switch may raise while reducing the weight for the lexical set for recognition before a switch according to the elapsed time to weight decision, after the above-mentioned weight decision section is required of a switch of the vocabulary for recognition by the above-mentioned lexical switch demand section for recognition. Therefore, a high recognition precision can be acquired also about the word of the lexical set for recognition before the above-mentioned switch.

[0061] Moreover, the voice recognition unit of invention of the above 1st computes the likelihood of each word which constitutes all the lexical sets for recognition for the above-mentioned recognition section. If it accomplishes so that the weight for the lexical set for recognition with which each word belongs to the value of the likelihood of each word may be applied and a word with the highest value may be made into a recognition result If the weight for the lexical set for recognition used for recognition and the weight for the lexical set for recognition which is not used for recognition are set up the optimal It can attain easily acquiring a high recognition precision, even when it utters with raising the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section, and the vocabulary for recognition before a speaker switching.

[0062] Moreover, the value of the weight for the lexical set for recognition corresponding to the content of an output which is outputting the voice recognition unit of invention of the above 1st when the lexical switch demand for recognition from the above-mentioned lexical switch demand section for recognition is made in the above-mentioned output section, Next, if a difference with the value of the weight for the lexical set for recognition corresponding to the content of an output which should be outputted becomes under a predetermined value and it will accomplish so that the above-mentioned content of an output may be switched In response to the above-mentioned lexical set for recognition being switched, the content of an output of the above-mentioned output section can be switched to the corresponding content of an output.

[0063] Moreover, the speech recognition approach of the 2nd invention determines the weight for two or more lexical sets for recognition which corresponded to the content of an output of the output section based on the time-of-day signal from the timer section. Since input voice is recognized using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made If the value of the weight for the lexical set for recognition before a switch is lowered in case the lexical set for recognition is switched, the recognition precision of the vocabulary for recognition after the switch according to the content of an output of the above-mentioned output section can be raised.

[0064] Furthermore, since a speaker recognizes it to it also using the word of the lexical set for recognition before a switch even if he utters to not knowing that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before a switch, a high recognition precision can be acquired also about the word of the lexical set for recognition before the above-mentioned switch.

[0065] Moreover, since the speech recognition processing program as which a computer is operated as the recognition section in claim 1, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section is recorded, the program documentation medium of the 3rd invention Like the case of claim 1, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section can be raised. Furthermore, even if a speaker utters that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before switching to not knowing, a high recognition precision can be acquired.

[Translation done.]

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

TECHNICAL FIELD

[Field of the Invention] This invention relates to the program documentation medium which recorded the speech recognition processing program on the voice recognition unit and the speech recognition approach of being carried in a computer or a Personal Digital Assistant and recognizing the voice by utterance of human being, and the list.

[Translation done.]

PRIOR ART

[Description of the Prior Art] In a voice recognition unit, in order to raise recognition precision, there is the recognition approach of switching the vocabulary for recognition if needed. As an application of the voice recognition unit using such a recognition approach, it is possible to perform the actuation guide of the device by the menu display using a display using speech recognition in the device which has displays, such as a personal computer and a Japanese word processor.

[0003] According to the above actuation guides, actuation can be studied, checking the display of the effectiveness by operating instructions or actuation in the pictures. And when there is little amount of information from the above-mentioned display -- the screen of the above-mentioned display is narrow -- the display of the actuation guide about two or more device actuation may be automatically switched with the passage of time. If voice is used for such an actuation guide, for a user, it is intelligible, and the number of manual operation buttons can be reduced and actuation can be simplified. In that case, if the vocabulary for recognition is switched with a switch of a display of the actuation guide about two or more device actuation, since the vocabulary for recognition can be lessened, a high recognition precision can be acquired.

[0004] In application of the recognition approach which switches such a vocabulary for recognition, two or more sets of the vocabulary for recognition which has relation in each menu which indicates by switch are memorized only several menu minutes. And by switching the vocabulary for recognition synchronizing with a switch of the menu display by actuation of a user, the passage of time, etc., in each menu, recognition processing can be performed for a necessary minimum vocabulary, and recognition precision can be raised. In that case, in case a menu display is automatically switched with the passage of time, a device will also switch the vocabulary for recognition automatically.

[0005] Hereafter, the voice recognition unit which can switch the above-mentioned vocabulary for recognition is explained. Drawing 4 is the block diagram showing an example of the voice recognition unit which can switch the above-mentioned vocabulary for recognition. Here, voice recognition unit 1 self shall perform automatically a switch of the content of a display according [this voice recognition unit 1] to a switch of the vocabulary for recognition, and the output section 13 for every predetermined time. A voice recognition unit 1 consists of the A/D (analog/digital) converter 2, the sonagraphy section 3, the recognition section 4, the sound model storing section 5, lexical storing [for recognition] / judgment section 6, the lexical identifier storage section 7 for the present recognition, the timer section 8, the lexical switch demand section 9 for recognition, the lexical switch demand time-of-day storage section 10 for recognition, a voice detecting element 11, the voice time-of-day storage section 12, and the output section 13.

[0006] The voice inputted into the above-mentioned voice recognition unit 1 is sent out and digitized by the A/D-conversion section 2 by the speaker. And this digitized voice wave is the sonagraphy section 3, and it is analyzed by the technique of the short-time analysis of a spectrum which shifts the above-mentioned time window every 8msec - 16msec while it hangs a short-time time window comparatively for every section of 20msec - 40msec. The voice wave started by the above-mentioned time window is changed into the time series of the feature vector of the unit called the frame which has the time amount length at the time of logging. Here, the above-mentioned feature vector is what extracted ****** of the voice spectrum in the time of day, it is usually 10-dimensional one - 100 dimensions, and the LPC (linear predictive coding) mel cepstrum multiplier etc. is used widely. In this way, the changed feature vector is outputted also to the voice detecting element 11 which detects initiation of voice input while it is sent out to the recognition section 4. If it does so, the voice time-of-day storage section 12 will detect and memorize the start time of voice input based on the voice input start signal from the voice detecting element 11, and the time-of-day signal from the timer section 8.

[0007] HMM (hidden Markov model) prepared for every recognition unit is prepared for the above-mentioned sound model storing section 5. Here, as the above-mentioned recognition unit, the phoneme and the word are used widely. Moreover, in HMM, it is the nondeterminism probability finite automaton which has two or more conditions, and is the source model of a statistical signal which expresses the source of an unsteady signal with connection of the source of a normal signal. In addition, parameters, such as a output probability and transition probability, are beforehand learned by the algorithm which gives corresponding study voice and is called a BAUMU-Welch algorithm. Hereafter, HMM whose recognition unit is a phoneme shall be memorized by the sound model storing section 5.

[0008] Actuation of a switch of the above-mentioned vocabulary for recognition applies the approach currently indicated by JP,6-337695,A. As the above-mentioned vocabulary for recognition, there shall be a lexical set A for recognition and a lexical set B for recognition, and the identifier of the lexical set A for recognition shall be memorized in this time by the lexical identifier storage section 7 for recognition. Moreover, the output section 13 shall show the content of a display corresponding to the lexical set A for recognition.

[0009] In this condition, progress of predetermined time makes advice from the timer section 8 to the lexical switch demand section 9 for recognition, and the output section 13. If it does so, the output section 13 will change the content of a display into the content of a display corresponding to the lexical set B for recognition. Moreover, a switch is required from the lexical switch demand section 9 for recognition, and the demand time of day is memorized by the lexical switch demand time-of-day storage section 10 for recognition. And the voice input start time Ts memorized by lexical storing [for recognition] / judgment section 6 at the demand time of day Tc and the voice time-of-day storage section 12 which are memorized by the lexical switch demand time-of-day storage section 10 for recognition is compared. When the voice input start time Ts is the

back [time of day / Tc / demand], since utterance was performed after the switch of the vocabulary for recognition was required, it is judged with the suitable lexical set for recognition being the lexical set B for recognition. It is judged with it being the lexical set A for recognition except it. And the content of storage of the lexical identifier storage section 7 for the present recognition is updated by the identifier of the corresponding lexical set for recognition.

[0010] After the judgment of the suitable lexical set for recognition is completed, in this way, the recognition section 4 The phoneme train of each word which constitutes which lexical set for recognition outputted from lexical storing [for recognition] / judgment section 6 corresponding to the feature vector obtained in the sonagraphy section 3, and the identifier memorized by the lexical identifier storage section 7 for the present recognition, Speech recognition is performed as follows using HMM stored in the sound model storing section 5.

[0011] That is, HMM of each word contained in the above-mentioned vocabulary for recognition is calculated first. HMM of each phoneme memorized by the sound model storing section 5 is made to specifically correspond to the phoneme train of each word which constitutes the lexical set for recognition, and it joins together.

[0012] Next, an occurrence probability is searched for about HMM of each word using the feature vector from the sonagraphy section 3. In the speech recognition by HMM, voice is expressed as time series of the symbol outputted from HMM between the state transitions from an initial state to a final state. Then, the probability for utterance to be generated from the model M (HMM of a word) can be searched for by setting the probability of an initial state to any value, and imposing a output probability and transition probability for every state transition one by one. On the contrary, when utterance is observed and it assumes that it generated from the model M with the utterance, the probability of generating from the model M can be calculated.

[0013] Hereafter, the recognition algorithm in the above-mentioned recognition section 4 is explained to a detail. The recognition section 4 considers the time series of the feature vector obtained by the sonagraphy section 3 as an input, searches for the occurrence probability about HMM of all the words contained in the vocabulary for recognition from lexical storing [for recognition] / judgment section 6, and makes a recognition result the word of HMM which presents the highest occurrence probability. Namely, the sequence of the input expressed by the time series of a feature vector is set to X=xvec1, xvec2 and xvec3, --, xvect, --, xvecI by making t (= 1, 2, --, I) into a frame number. In addition, "xveci" is the vector of many dimensions. Hereafter, "xvec" is written. [Vector x] Furthermore, the set of the initial state of Model M is set to S, and the set of a final state is set to F. Moreover, the sere of the j-th condition is expressed as Q=qOj, q1j, q2j, --, qtj, --, qIj by making "i, j" into a state number. In a top type, "qtj" expresses the condition of having changed by the input symbol xvect of the t-th frame. Here, it is qOj**S and qIj**F. Furthermore, when the initial probability of an initial state is expressed with pi i:sigma qi**Spii =1, transition probability from Condition qi to Condition qj is set to aij and the output probability by which xveci is then outputted is set to bij (xveci), occurrence probability (likelihood) P(X|M) of an input sequence is,

$$P(X \mid M) = \sum_{allQj} \pi_0^j \prod_{i=1}^{I} a_{t-1,t}^j \cdot b_{t-1,t}^j (x_{vec} i)$$

It is come out and expressed. It calculates by HMM-attaching, and outputs and displays on the output section 13 by making into a recognition result the word corresponding to HMM corresponding to all the words contained in the vocabulary for recognition in the operation of this occurrence probability (likelihood) P(X|M) which presents the highest occurrence probability (likelihood) P.

[Translation done.]

EFFECT OF THE INVENTION

[Effect of the Invention] As mentioned above, it is the voice recognition unit of the 1st invention so that clearly, Two or more lexical sets for recognition corresponding to the content of an output of the output section are stored in the recognition lexical storing section. By the weight decision section Since the weight for each above-mentioned lexical set for recognition is determined based on the time-of-day signal from the timer section and input voice is recognized by the recognition section using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned] It is based on the switch demand of the vocabulary for recognition by the lexical switch demand section for recognition. If the value of the weight for the lexical set for recognition before a switch is lowered in case it is switched to the lexical set for recognition according to a switch of the content of an output of the above-mentioned output section, the recognition precision of the vocabulary for recognition after a switch can be raised.

[0058] Furthermore, since a speaker recognizes it to it also using the word of the lexical set for recognition before a switch even if he utters to not knowing that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before a switch, a high recognition precision can be acquired also about the word of the lexical set for recognition before the above-mentioned switch.

[0059] That is, according to this invention, even when switching the vocabulary for recognition automatically, a high recognition precision can be acquired. Furthermore, the voice recognition unit which a speaker is not made to pay a certain actuation and latency time in that case, and is easy to use is realizable.

[0060] Moreover, the voice recognition unit [above / 1st] of invention can switch gradually the above-mentioned vocabulary for recognition used for recognition, if it accomplishes so that the weight for the lexical set for recognition after a switch may raise while reducing the weight for the lexical set for recognition before a switch according to the elapsed time to weight decision, after the above-mentioned weight decision section is required of a switch of the vocabulary for recognition by the above-mentioned lexical switch demand section for recognition. Therefore, a high recognition precision can be acquired also about the word of the lexical set for recognition before the above-mentioned switch.

[0061] Moreover, the voice recognition unit of invention of the above 1st is the above-mentioned recognition section. If it accomplishes so that the likelihood of each word which constitutes all the lexical sets for recognition may be computed, the weight for the lexical set for recognition with which each word belongs to the value of the likelihood of each word may be applied and a word with the highest value may be made into a recognition result If the weight for the lexical set for recognition used for recognition and the weight for the lexical set for recognition which is not used for recognition are set up the optimal It can attain easily acquiring a high recognition precision, even when it utters with raising the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section, and the vocabulary for recognition before a speaker switching.

[0062] Moreover, the voice recognition unit of invention of the above 1st is the value of the weight for the lexical set for recognition corresponding to the content of an output which is outputting the above-mentioned output section when the lexical switch demand for recognition from the above-mentioned lexical switch demand section for recognition is made, Next, if a difference with the value of the weight for the lexical set for recognition corresponding to the content of an output which should be outputted becomes under a predetermined value and it will accomplish so that the above-mentioned content of an output may be switched, in response to the above-mentioned lexical set for recognition being switched, the content of an output of the above-mentioned output section can be switched to the corresponding content of an output.

[0063] moreover -- since the speech recognition approach of the 2nd invention determines the weight for two or more lexical sets for recognition which corresponded to the content of an output of the output section based on the time-of-day signal from the timer section and input voice is recognized using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made If the value of the weight for the lexical set for recognition before a switch is lowered in case the lexical set for recognition is switched, the recognition precision of the vocabulary for recognition after the switch according to the content of an output of the above-mentioned output section can be raised.

[0064] Furthermore, since a speaker recognizes it to it also using the word of the lexical set for recognition before a switch even if he utters to not knowing that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before a switch, a high recognition precision can be acquired also about the word of the lexical set for recognition before the above-mentioned switch.

[0065] Moreover, the program documentation medium of the 3rd invention is since the speech recognition processing program as which a computer is operated as the recognition section in claim 1, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section is recorded, Like the case of claim 1, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section can be raised. Furthermore, even if a speaker utters that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before switching to not knowing, a high recognition precision can be acquired.

[Translation done.]

TECHNICAL PROBLEM

[Problem(s) to be Solved by the Invention] However, there are the following problems in the voice recognition unit which applied the lexical switch actuation for recognition indicated by above-mentioned conventional JP,6-337695,A. That is, as mentioned above, when the voice input start time Ts is the back [time of day / Tc / for recognition / lexical switch demand] in the lexical switch actuation for recognition indicated by JP,6-337695,A, he is trying to switch the set of the vocabulary for recognition. When the lexical switch demand for recognition is made by actuation of a speaker, since utterance is performed after the switch demand of the vocabulary for recognition is surely made, this approach is effective.
[0015] However, like the voice recognition unit shown in drawing 4 , when it is the voice recognition unit which the vocabulary for recognition turns off and is automatically replaced with the passage of time, a switch of the vocabulary for recognition is performed regardless of a speaker's consciousness at all. Therefore, a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and when a switch of the vocabulary for recognition is performed automatically, the need of returning to the established state of the vocabulary for recognition before the switch which the speaker wanted to utter by a certain approach arises. And there is a problem that a speaker will be kept waiting until it makes a speaker pay a certain actuation or the vocabulary for recognition before a switch is set up automatically in that case.
[0016] Then, the object of this invention is to offer the program documentation medium which recorded the speech recognition processing program on the voice recognition unit which is easy to use and the speech recognition approach that a high recognition precision is acquired, and the list, even when switching the vocabulary for recognition automatically.
[0017]

[Translation done.]

MEANS

[Means for Solving the Problem] The recognition section which recognizes the voice into which the 1st invention was inputted in order to attain the above-mentioned object, The output section which outputs information including the recognition result of this recognition section, and the recognition lexical storing section in which the vocabulary for recognition used at the time of the above-mentioned recognition was stored, In the voice recognition unit which has the timer section and the lexical switch demand section for recognition which requires a switch of the above-mentioned vocabulary for recognition based on the time-of-day signal from this timer section the above-mentioned output section The switch output of two or more contents of an output is carried out. The above-mentioned vocabulary for recognition It is classified into two or more lexical sets for recognition which become by the set of the word for recognition corresponding to the content of an output of the above-mentioned output section, and a switch of the above-mentioned vocabulary for recognition is performed in the unit of the above-mentioned lexical set for recognition. Based on the time-of-day signal from the above-mentioned timer section, it has the weight decision section which determines the weight for each above-mentioned lexical set for recognition, and the above-mentioned recognition section is characterized by recognizing input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].

[0018] According to the above-mentioned configuration, input voice is recognized by the recognition section using the weight for each lexical set for recognition determined by the weight decision section based on the time-of-day signal from all the lexical sets for recognition, and the timer section. If a switch of the vocabulary for recognition is required by the lexical switch demand section for recognition based on the time-of-day signal from the above-mentioned timer section in that case, the lexical set for recognition used now will be switched to the lexical set for recognition according to a switch of the content of an output of the output section. Therefore, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised.

[0019] Furthermore, since recognition is performed also using the word of the lexical set for recognition before a switch even if it utters with the vocabulary for recognition before a switch to not knowing, a high recognition precision is acquired [that the above-mentioned lexical set for recognition was switched for the speaker, and] also about the word of the lexical set for recognition before the above-mentioned switch.

[0020] Moreover, after the above-mentioned weight decision section is required of a switch of the vocabulary for recognition by the above-mentioned lexical switch demand section for recognition, while the voice recognition unit [above / 1st] of invention reduces the weight for the lexical set for recognition before a switch according to the elapsed time to weight decision, it is [voice recognition unit] desirable in accomplishing so that the weight for the lexical set for recognition after a switch may be raised.

[0021] While according to the above-mentioned configuration the elapsed time after a switch of the vocabulary for recognition is required by the above-mentioned lexical switch demand section for recognition takes for becoming long and the recognition precision of the vocabulary for recognition before a switch becomes low, the recognition precision of the vocabulary for recognition after a switch becomes high. In this way, a switch of the above-mentioned vocabulary for recognition used for recognition is performed gradually.

[0022] Moreover, as for the voice recognition unit of invention of the above 1st, it is desirable to accomplish so that the likelihood of each word which constitutes all the above-mentioned lexical sets for recognition for the above-mentioned recognition section may be computed, the weight for the lexical set for recognition with which each word belongs to the value of the likelihood of each word may be applied and a word with the highest value may be made into a recognition result.

[0023] Obtaining raising the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section and a recognition precision high even when a speaker utters at the vocabulary for recognition before switching is easily attained by setting up the weight for the lexical set for recognition used for recognition, and the weight for the lexical set for recognition which is not used for recognition the optimal according to the above-mentioned configuration.

[0024] Moreover, when the difference of the value of the weight for the lexical set for recognition corresponding to the content of an output which is outputting the above-mentioned output section when the lexical switch demand for recognition from the above-mentioned lexical switch demand section for recognition is made, and the value of the weight for the lexical set for recognition corresponding to the content of an output which should output to a degree becomes under in a predetermined value, it is [voice recognition unit / of invention of the above 1st] desirable in accomplishing so that the above-mentioned content of an output may switch.

[0025] According to the above-mentioned configuration, in response to the above-mentioned lexical set for recognition being switched, it is switched to the content of an output to which the content of an output of the above-mentioned output section is equivalent.

[0026] Moreover, face the speech recognition approach of the 2nd invention recognizing the inputted voice using the vocabulary for recognition, and outputting a recognition result, and it is set to the speech recognition approach which switches the above-mentioned vocabulary for recognition automatically based on the time-of-day signal from the timer section. Carry

out the switch output of two or more contents of an output at the output section, and the above-mentioned vocabulary for recognition is switched in the unit of two or more lexical sets for recognition which become by the set of the word for recognition corresponding to each above-mentioned content of an output. Based on the time-of-day signal from the above-mentioned timer section, the weight for each above-mentioned lexical set for recognition is determined, and it is characterized by recognizing the above-mentioned input voice using each weight by which a vocabularies [all / for recognition] set and the above-mentioned decision were made [above-mentioned].

[0027] According to the above-mentioned configuration, input voice is recognized using the weight for each lexical set for recognition determined based on the time-of-day signal from all the lexical sets for recognition, and the timer section. If a switch of the vocabulary for recognition is required based on the time-of-day signal from the above-mentioned timer section in that case, the lexical set for recognition used now will be switched to the lexical set for recognition according to a switch of the content of an output of the output section. Therefore, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised.

[0028] Furthermore, since recognition is performed also using the word of the lexical set for recognition before a switch even if it utters with the vocabulary for recognition before a switch to not knowing, a high recognition precision is acquired [that the above-mentioned lexical set for recognition was switched for the speaker, and] also about the word of the lexical set for recognition before the above-mentioned switch.

[0029] Moreover, the program documentation medium of the 3rd invention is characterized by recording the speech recognition processing program as which a computer is operated as the recognition section in claim 1, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section.

[0030] According to the above-mentioned configuration, like the case of claim 1, if the value of the weight for the lexical set for recognition before a switch is lowered, the recognition precision of the vocabulary for recognition after the switch corresponding to the content of an output of the above-mentioned output section will be raised. Furthermore, even if a speaker utters that the above-mentioned lexical set for recognition was switched with the vocabulary for recognition before switching to not knowing, a high recognition precision is acquired.

[0031]

[Embodiment of the Invention] Hereafter, the gestalt of implementation of a graphic display explains this invention to a detail. Drawing 1 is a block diagram in the voice recognition unit of the gestalt of this operation. This voice recognition unit 21 consists of the voice input section 22, the A/D-conversion section 23, the sonagraphy section 24, the recognition section 25, the sound model storing section 26, the lexical set storing section 27 for the 1st recognition, the lexical set storing section 28 for the 2nd recognition, the weighting-factor decision section 29, the timer section 30, the lexical switch demand section 31 for recognition, and the output section 32.

[0032] The above-mentioned voice input section 22 is equipped with the audio input unit containing a microphone, changes the inputted voice into an electrical signal (sound signal), and outputs it to the A/D-conversion section 23. The A/D-conversion section 23 changes the sound signal which is an inputted analog signal into a digital signal, and outputs the digitized sound signal to the sonagraphy section 24. In addition, the sound signal by which digitization was carried out [above-mentioned] is expressed with the time series of amplitude value.

[0033] The above-mentioned sonagraphy section 24 extracts a feature vector from the digital sound signal from the A/D-conversion section 23 for every frame, and outputs it to the recognition section 25. Here, the above-mentioned feature vector continues and arranges the 34-dimensional vector xvec which consists of power of the power of the sound signal in each frame, primary - the 16th LPC cepstrum multiplier, and a before frame, and a total of 34 elements of the LPC cepstrum multiplier (primary - 16th order) of a before frame on all frames (t= 1, 2, --, I).

[0034] The above-mentioned recognition section 25 calculates the occurrence probability (likelihood) P of each word which constitutes the lexical set B for recognition stored in the lexical set A for recognition stored in the lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition using a sound model using the technique explained by the Prior art using the feature vector extracted in the sonagraphy section 24. Furthermore, weight w determined in the weighting-factor decision section 29 is hung on the likelihood P of each word, and the word corresponding to HMM which presents highest likelihood w-P is outputted to the output section 32.

[0035] The sound model used in case the above-mentioned sound model storing section 26 performs speech recognition in the recognition section 25 is stored. The above-mentioned sound model makes a phoneme a unit, and HMM learned by the algorithm beforehand called a BAUMU-Welch algorithm using the study voice of an unspecified speaker (initial study) is used. In addition, Above HMM is memorized in the array for several condition minutes which uses the transition probability and output probability distribution in each condition as an element. Moreover, the above-mentioned transition probability is memorized in the array for several transition minutes by using transition probability to each condition as an element. Moreover, the above-mentioned output probability is expressed with the contaminated normal distribution of the many dimensions which carried out weighting addition of two or more normal distribution, and is memorized in the array for the number of dimension which uses the weight, the mixed mean vector, and mixed distributed vector in each normal distribution as an element. Here, it is expressed with the array of the element of "34" as the number of elements of the feature vector extracted from a sound signal for every frame in the sonagraphy section 24 with same above-mentioned mean vector and distributed vector.

[0036] The above-mentioned timer section 30 outputs the time-of-day signal showing time of day to the lexical switch demand section 31 for recognition, the weighting-factor decision section 29, and the output section 32, and notifies time of day. If it does so, the lexical switch demand section 31 for recognition will judge whether a switch of the vocabulary for recognition is required based on the time of day by which advice was given [above-mentioned]. And in requiring, it requires a switch of the vocabulary for recognition from the weighting-factor decision section 29.

[0037] The inside of the lexical set B for recognition stored in the lexical set A for recognition with which the above-mentioned weighting-factor decision section 29 is stored in the lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition, The weight w2 applied to the word which constitutes the lexical set for recognition

corresponding to the content of a display in which it is indicated by current by the output section 32, and the weight w1 applied to the word which constitutes the lexical set for recognition corresponding to the content of a display which is not shown by the output section 32 are determined. Such weight w1 and w2 is determined whenever predetermined time deltaT passes using the weight function W1 (t) memorized and W2 (t) on the basis of the time of day tO from the timer section 30 when a switch is required from the lexical switch demand section 31 for recognition. And the sequential output of the value of both the determined weight w1 and w2 is carried out at the recognition section 25.

[0038] The word which constitutes each lexical set A and B for recognition is memorized by the above-mentioned lexical set storing section 27 for the 1st recognition, and the lexical set storing section 28 for the 2nd recognition in the array for several alphabetic character minutes which uses the character string of the notation of each word, and a phoneme train as an element.

[0039] The above-mentioned output section 32 is equipped with an image display device including a display, and stores the content of the 1st display corresponding to the lexical set A for recognition, and the content of the 2nd display corresponding to the lexical set B for recognition. And it judges whether based on the time of day notified from the timer section 30, the content of a display which is indicating by current among the contents of the 1st and 2nd display is changed, and when changing, the content of a display of a screen is switched. Furthermore, the recognition result from the recognition section 25 is displayed on a screen.

[0040] Drawing 2 shows time amount change with the weight function W2 for the lexical set for recognition corresponding to the content of a display which the above-mentioned output section 32 has chosen now (t), and the weight function W1 for the lexical set for recognition corresponding to the content of a non-choosing display (t). It is begun from about zero predetermined value "a" smaller than 1 to carry out the monotonous increment of the value of a weight function W1 (t) at the time of day t0 when the switch demand of the vocabulary for recognition was outputted, and becomes a value "1" after time of day t2. On the other hand, it is begun at time of day t0 to carry out monotonous reduction of the value of a weight function W2 (t) with the value of a weight function W1 (t) at reverse from a value "1", and becomes a predetermined value "a" after time of day t2. In that case, the difference of weight w1 and weight w2 serves as a threshold h at time of day t1. And the output section 32 will switch the content of a display currently displayed on the screen, if the value of this difference turns into under the threshold h (i.e., if time amount T (> (t1-t0)) passes since the time of day t0 when the switch of the vocabulary for recognition was demanded).

[0041] That is, when judging that the above-mentioned output section 32 changes the content of a display based on the time of day notified from the timer section 30, it is set up so that only the above-mentioned time amount T may be late for the event of the lexical switch demand section 31 for recognition judging that the above-mentioned switch is required based on the time of day notified from the timer section 30.

[0042] Thus, in the gestalt of this operation, although the content of a display of a screen is automatically switched by the output section 32, even if it is after cutting and replacing, even if it is before the content of a display cuts and replaces, the recognition section 25 calculates likelihood P for the vocabulary of the lexical set A for recognition, and both the lexical set of the lexical set B for recognition. And if it is before the content switch of a display, it is 1>w> (1+a+h)/2, and if it is after a switch, weight w which is 1>w> (1+a-h)/2 will be hung on the likelihood P of the word which constitutes the lexical set for recognition corresponding to the content of a display chosen by the current output section 32. On the other hand, if it is before the content switch of a display (1+a-h), it is /2>w>a, and if it is after a switch (1+a+h), weight w which is /2>w>a will be hung on the likelihood P of the word which constitutes the lexical set for recognition corresponding to the content of a display by the side of un-choosing. In this way, he calculates final likelihood w-P and is trying to determine a recognition result.

[0043] If it puts in another way, a switch of the vocabulary for recognition in the conventional voice recognition unit shown in drawing 4 It sets in the gestalt of this operation to carrying out by switching the vocabulary for recognition itself used for the operation of likelihood P. The two-set vocabulary for recognition used for the operation of likelihood P is performed by changing gradually the value of weight w hung on likelihood P, without switching between "1" and about zero predetermined value "a."

[0044] Therefore, in the gestalt of this operation, even if a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and count of likelihood w-P about the word of the vocabulary for recognition before a switch will also be performed also after a switch of the vocabulary for recognition is performed automatically, and it utters with the vocabulary for recognition before a speaker switching, it becomes possible to recognize correctly. Moreover, the function which raises the recognition precision of the vocabulary corresponding to the content of a display of the output section 32 is not spoiled like the case where the vocabulary for recognition itself is switched like the voice recognition unit shown in drawing 4 in that case.

[0045] Drawing 3 is the flow chart of the weight decision processing actuation performed by the above-mentioned weighting-factor decision section 29. Hereafter, actuation of weight decision is explained according to drawing 3 . Here, the output section 32 sets to W2 (t) the weight function for the lexical set for recognition corresponding to the content of a display which is making current selection, and sets the weight function for the lexical set for recognition corresponding to the content of a non-choosing display to W1 (t). If a switch is required from the lexical switch demand section 31 for recognition, weight decision processing actuation will start.

[0046] At step S1, the switch demand time of day t0 of the vocabulary for recognition is acquired based on the time-of-day signal from the above-mentioned timer section 30. At step S2, the count j of calculation of the weight value w is initialized by "0." At step S3, the increment of the count j of calculation is carried out. By step S4, since the switch demand time of day t0 is acquired, or after computing the weight value w last time, it is distinguished whether predetermined time deltaT passed. Consequently, if it has passed, it will progress to step S5. At step S5, it is distinguished whether current time of day (t0+j-deltaT) is over time of day t2. Consequently, if it has exceeded, while ending weight decision processing actuation, if it has not exceeded, it progresses to step S6.

[0047] At step S6, the function number i of the above-mentioned weight function Wi (t) is initialized by "1." At step S7, the value wi of weight is computed by "j-deltaT" being substituted for elapsed time t from the switch demand time of day t0 in a

weight function W1 (t). The increment of the function number i is carried out at step S8. By step S9, it is distinguished whether the value of the function number i is larger than "2." Consequently, if larger than "2" while with "2" carrying out a return to step S7 and shifting to calculation of the weight value w2, it will be judged that the weight in the present time of day corresponding to all the lexical sets A and B for recognition was computed, and it will progress to step S10. [below] At step S10, the array of the weight values w1 and w2 in the present time of day by which calculation was carried out [above-mentioned] is outputted to the recognition section 25. A return is carried out to step S3 such the back, and it shifts to calculation of the weight values w1 and w2 in the next time of day.

[0048] Henceforth, if the above-mentioned step S3 - step S10 were repeated, current time of day is over time of day t2 in step S5 and it will be distinguished, weight decision processing actuation will be ended. After that, "1" is outputted to every predetermined time deltaT as a weight value w2 for the lexical set for recognition corresponding to the content of a display, and a predetermined value "a" is outputted to every predetermined time deltaT as a weight value w1 for the lexical set for recognition corresponding to the content of a non-choosing display. And if a switch demand is outputted from the lexical switch demand section 31 for recognition next, the above-mentioned weight decision processing actuation will start.

[0049] As mentioned above, the recognition section 25 in the gestalt of this operation computes the likelihood P of the word which constitutes the lexical set A for recognition stored in the lexical set storing section 27 for the 1st recognition, and the lexical set B for recognition stored in the lexical set storing section 28 for the 2nd recognition using the sound model stored in the sound model storing section 26. A switch of the lexical set for recognition accompanying a switch of the content of a display of the output section 32 in that case is performed by switching the value of the weight w2 and w1 applied to the likelihood P of the word which constitutes selection and the lexical set for non-choosing recognition to "1" and about zero predetermined value "a" rather than switching the lexical set for recognition itself. And in that case, in proportion to the elapsed time "j-deltaT" from the time of day t0 when the switch demand was made from the lexical switch demand section 31 for recognition, he value "a" Passes from a value "1" gradually rather than switching the value of weight w2 and w1 gradually, or is trying to switch to a value "1" from a value "a."

[0050] Therefore, since according to the gestalt of this operation count of likelihood w-P about the word of the lexical set for recognition before a switch is also performed even if the vocabulary for recognition will be switched automatically [a speaker misses the opportunity of utterance of the vocabulary for recognition by a certain reason, and], even if it utters with the vocabulary for recognition before a speaker switching, it can recognize correctly. Moreover, the function which raises the recognition precision of the vocabulary for recognition corresponding to the content of a display of the output section 32 is not spoiled like the case where the vocabulary for recognition itself is switched in that case like the voice recognition unit shown in drawing 4 .

[0051] In addition, he is trying to switch linearly the weight function W2 for the lexical set for selection recognition (t), and the weight function W1 for the lexical set for recognition corresponding to the content of a non-choosing display (t) to a value "a", and "1" from a value "1", and "a" in the gestalt of the above-mentioned implementation in proportion to the elapsed time "j-deltaT" from the switch demand time of day t0 by the lexical switch demand section 31 for recognition. However, in this invention, a function W1 (t) and the configuration of W2 (t) are not limited to a straight line. While making it a curve, lowering the value of a function W1 (t) while raising the value of the function W2 by the switch time of day t1 of the content of a display (t), and lowering the value of the function W2 after the switch time of day t1 of the content of a display (t), the value of a function W1 (t) may be raised.

[0052] Moreover, whenever predetermined-time deltaT passes the above-mentioned weighting-factor decision section 29 on the basis of the switch demand time of day tO from the lexical switch demand section 31 for recognition, it constitutes so that the weight values w1 and w2 may be determined and it may output to the recognition section 25, and in the gestalt of the above-mentioned implementation, the recognition section 25 constitutes so that the weight values w1 and w2 inputted may be used if needed and recognition processing may be carried out. However, this invention is not limited to this, in case it recognizes the recognition section 25, it is constituted so that a weight decision demand may be advanced to the weighting-factor decision section 29, and if a weight decision demand is received, the weighting-factor decision section 29 will not interfere, even if it constitutes so that the elapsed time from the switch demand time of day tO by the lexical switch demand section 31 for recognition may be substituted and computed to a weight function Wi (t).

[0053] By the way, the function as the above-mentioned recognition section in a gestalt, the output section, the timer section, the lexical switch demand section for recognition, and the weight decision section of each above-mentioned implementation is realized by the speech recognition processing program recorded on the program documentation medium. The above-mentioned program documentation media in the gestalt of the above-mentioned implementation are program media which become by ROM (read only memory). Or you may be the program media by which reading appearance is equipped with and carried out to external auxiliary storage. In addition, the program read-out means which reads a speech recognition processing program from the above-mentioned program media in the case of which may have the configuration which carries out direct access to the above-mentioned program media and which is read to them, and may download it to the program storage area (not shown) prepared in RAM (random access memory), and you may have the configuration accessed and read to the above-mentioned program storage area. In addition, the download program for downloading from the above-mentioned program media to the above-mentioned program storage area of RAM shall be beforehand stored in the main frame.

[0054] With the above-mentioned program media, it is constituted disengageable a body side here. Magnetic disks, such as tape systems, such as a magnetic tape and a cassette tape, a floppy (trademark) disk, and a hard disk, CD(compact disk)-ROM and MO (optical MAG) disk, MD (mini disc), The disk system of optical disks, such as DVD (digital video disc), It is the medium including semiconductor memory systems, such as card systems, such as IC (integrated circuit) card and an optical card, a mask ROM, EPROM (ultraviolet-rays elimination mold ROM) and EEPROM (electric elimination mold ROM), and a flash ROM, which **** a program fixed.

[0055] Moreover, if it has the configuration which the voice recognition unit in the gestalt of each above-mentioned implementation is equipped with a modem, and contains the Internet and in which a communication network and connection are possible, even if the above-mentioned program media are media which **** a program fluidly by download from a

communication network etc., they will not interfere. In addition, the download program for downloading from the above-mentioned communication network which can be set in that case shall be beforehand stored in the main frame. Or it shall be installed from another record medium.

[0056] In addition, it is not limited only to a program and what is recorded on the above-mentioned record medium can also record data.

[Translation done.]

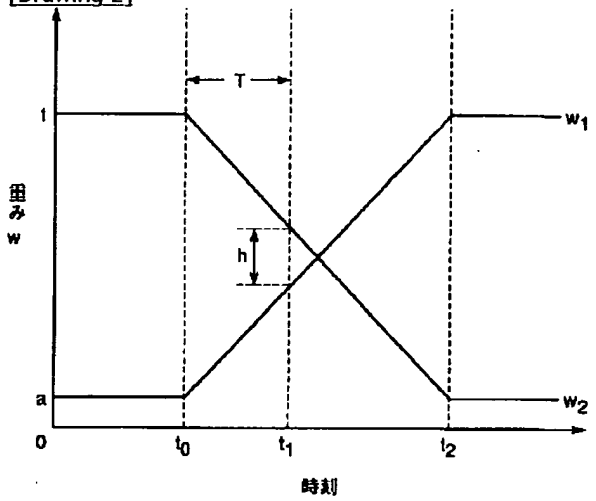
DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is a block diagram in the voice recognition unit of this invention.
[Drawing 2] It is drawing showing selection and time amount change of the weight function for the lexical set for non-choosing recognition.
[Drawing 3] It is the flow chart of the weight decision processing actuation performed by the weighting-factor decision section in drawing 1 .
[Drawing 4] It is the block diagram of the conventional voice recognition unit which can switch the vocabulary for recognition.
[Description of Notations]
21 -- Voice recognition unit,
22 -- Voice input section,
23 -- A/D-conversion section,
24 -- Sonagraphy section,
25 -- Recognition section,
26 -- Sound model storing section,
27 -- Lexical set storing section for the 1st recognition,
28 -- Lexical set storing section for the 2nd recognition,
29 -- Weighting-factor decision section,
30 -- Timer section,
31 -- Lexical switch demand section for recognition,
32 -- Output section.

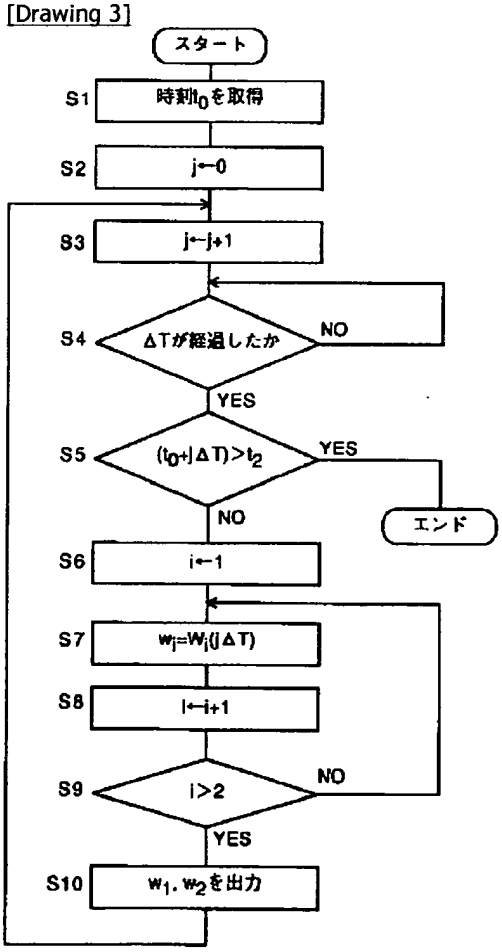
[Translation done.]

DRAWINGS

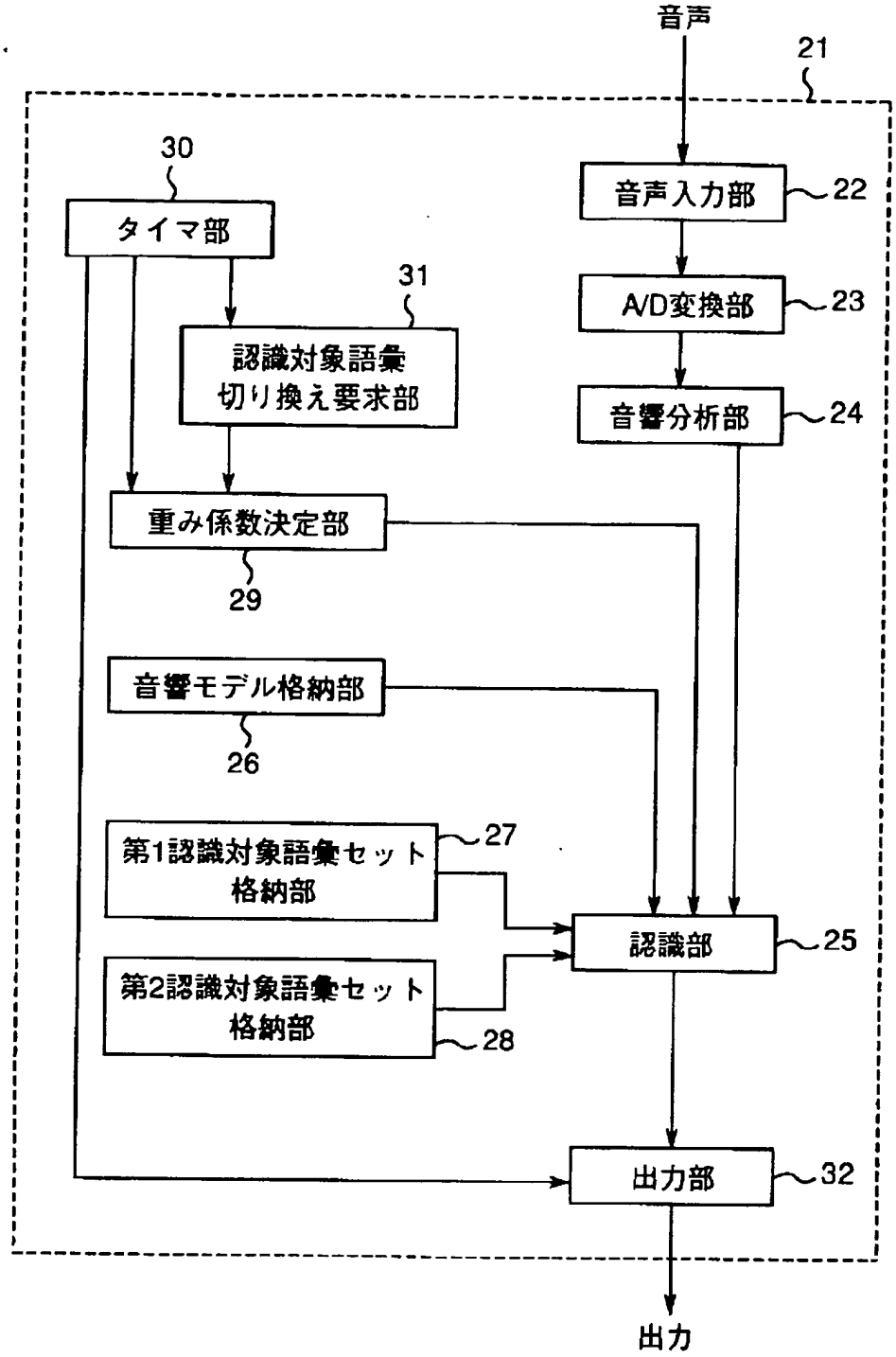
[Drawing 2]



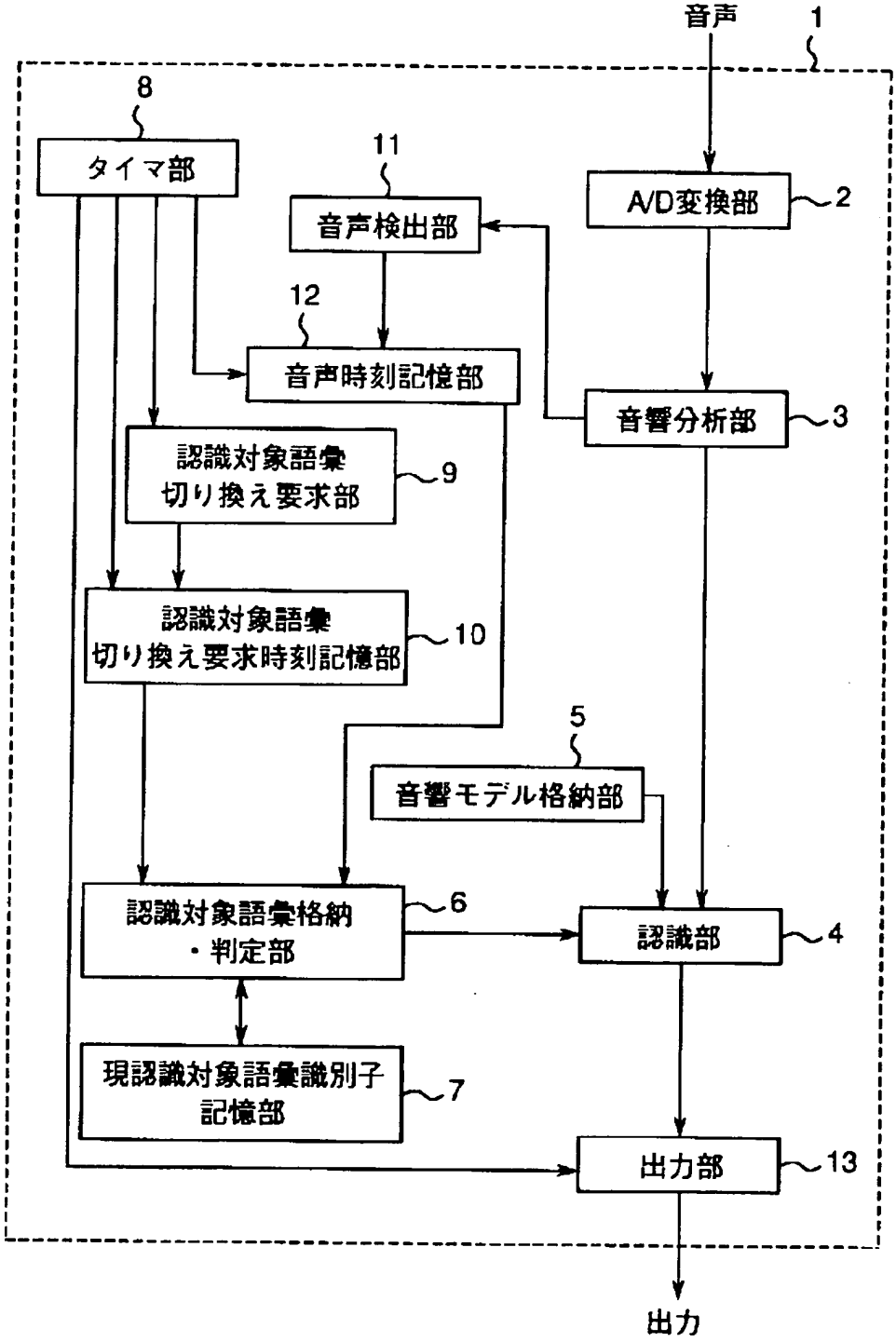
[Drawing 3]



[Drawing 1]


音声
21
30
タイマ部
音声入力部
22
31
認識対象語彙
切り換え要求部
A/D変換部
23
音響分析部
24
重み係数決定部
29
音響モデル格納部
26
第1認識対象語彙セット
格納部
27
認識部
25
第2認識対象語彙セット
格納部
28
出力部
32
出力

[Drawing 4]



[Translation done.]